



L'orientation scolaire et professionnelle

31/1 | 2002
Varia

Intérêts et apports de l'analyse des variables latentes pour le chercheur en sciences sociales : exemple d'application à l'économie de l'éducation

The contribution of the linear structural relation analysis within the context of social sciences

Sophie Morlaix



Electronic version

URL: <http://journals.openedition.org/osp/4880>

DOI: 10.4000/osp.4880

ISSN: 2104-3795

Publisher

Institut national d'étude du travail et d'orientation professionnelle (INETOP)

Printed version

Date of publication: 15 March 2002

ISSN: 0249-6739

Electronic reference

Sophie Morlaix, « Intérêts et apports de l'analyse des variables latentes pour le chercheur en sciences sociales : exemple d'application à l'économie de l'éducation », *L'orientation scolaire et professionnelle* [Online], 31/1 | 2002, Online since 01 March 2005, connection on 30 April 2019. URL : <http://journals.openedition.org/osp/4880> ; DOI : 10.4000/osp.4880

This text was automatically generated on 30 April 2019.

© Tous droits réservés

Intérêts et apports de l'analyse des variables latentes pour le chercheur en sciences sociales : exemple d'application à l'économie de l'éducation

The contribution of the linear structural relation analysis within the context of social sciences

Sophie Morlaix

- 1 L'un des objectifs du chercheur en sciences sociales consiste à réduire la complexité des faits observés de façon à dégager des relations entre plusieurs variables caractéristiques et explicatives du phénomène étudié. En fonction d'une part des données dont il dispose, et d'autre part de ses objectifs, il lui est possible d'arbitrer entre différentes méthodes économétriques. Parmi les outils disponibles, certaines méthodes d'estimation restent peu utilisées compte tenu de leur intérêt parfois méconnu car d'un abord difficile. Parmi celles-ci figure l'analyse en structure de covariances (modèles LISREL notamment).
- 2 L'ambition de cet article vise à présenter une explication simplifiée de l'apport de ces modèles, dans les sciences sociales en général. Dans un souci de lisibilité et de clarté, il est également proposé de compléter la présentation générique de ce type de modèle par un exemple concret d'application. Dans ce sens, est exposée une analyse plus spécifique au niveau du fonctionnement du système éducatif.

Présentation des modèles en structures de covariances

Généralités

- 3 D'après Aish-Van Vaerenbergh (1997), l'analyse en structure de covariances ¹ résulte de la convergence des méthodes d'analyses développées dans quatre disciplines relativement indépendantes : « *l'analyse factorielle en psychométrie* (introduite par Spearman, 1904), *la path analysis en biométrie* (introduite par Wright, 1934), *et ultérieurement en sociologie* (Blalock, 1961, Boudon, 1965, Duncan, 1966), *et l'analyse des systèmes d'équations simultanées en économétrie* (Golberger, 1971) ». L'intégration de ces méthodes est due à Jöreskog (1967) qui développe l'analyse en structure de covariances, initiant ainsi les modèles LISREL ², et le logiciel du même nom (Jöreskog & Van Thillo, 1973 ; Jöreskog & Sörbom, 1982).
- 4 Cinq spécificités principales distinguent l'analyse en structure de covariances, et lui permettent de faire partie de ce que certains auteurs, tel Valette-Florence (1988), qualifieront de « *méthodes d'analyse multi variée de la deuxième génération* ³ ». Ses caractéristiques sont exposées brièvement par la suite et seront reprises, de façon plus détaillée, tout au long de la présentation des modèles LISREL. Ces méthodes d'analyse multi variée, dites de la seconde génération, permettent ainsi conjointement :
 - 5 d'introduire dans l'analyse, des variables latentes (ou non observables),
 - 6 de spécifier la nature des relations entre ces variables latentes et leurs mesures,
 - 7 d'offrir la possibilité de préciser le type de relations envisagées entre variables latentes,
 - 8 d'analyser les inférences causales entre plusieurs ensembles de variables explicatives et expliquées,
 - 9 de se placer dans une démarche confirmative.
- 10 L'un des principaux apports des modèles LISREL réside dans la prise en compte de variables latentes dans l'explication de relations entre plusieurs variables. L'une de ses particularités essentielles est d'expliquer une matrice de variances-covariances entre des variables que l'on aurait observées, en formulant l'hypothèse d'existence d'une structure causale sous-jacente à l'ensemble du modèle, et en supposant alors que les corrélations entre variables observées (indicateurs) permet de rendre explicite une dimension latente des phénomènes, en postulant l'existence de variables inobservées (ou inobservables) qui rendent compte des relations entre variables observées.
- 11 La prise en compte d'une démarche du type LISREL contraint à la formulation d'un modèle théorique postulé *a priori* dont le test va permettre de comprendre dans quelle mesure un modèle théorique rend compte des relations (corrélations) entre les variables observées. « *L'analyse des structures de covariances vise à vérifier la validité d'une théorie causale préalablement formulée en testant l'ajustement d'un modèle mathématique à des données observées* » (Aish-Van Vaerenbergh, 1997). Le modèle LISREL repose sur une analyse confirmative. Il relève d'une démarche hypothético-déductive dans laquelle on formule des hypothèses *a priori* que l'on va confronter avec les faits.
- 12 Soumettre une structure causale hypothétique à l'épreuve de l'observation, implique plusieurs étapes, dans la mesure notamment où LISREL se décompose en deux modèles : le

modèle de mesure d'une part, qui établit les relations entre les concepts (variables latentes) et les indicateurs de ces concepts (variables observées dont on dispose d'une mesure) et le modèle structurel d'autre part, qui détermine les liaisons entre les différentes variables latentes (concepts). Les étapes et conditions nécessaires au test d'une structure causale sous LISREL sont explicitées par la suite. Le principal objectif de cette description n'est toutefois pas de montrer la complexité des phénomènes qui peuvent être appréhendés par des méthodes statistiques telles que LISREL. Elle vise plutôt l'exposé de quelques notions simples permettant de comprendre ou rappeler les éléments fondamentaux qui constituent la méthodologie de ces modèles.

Les différentes étapes du test d'une structure causale hypothétique

Le développement d'un modèle théorique

L'utilité d'une théorie principale (le modèle structural)

- 13 La théorie principale est à la base du modèle. Elle sert à développer un modèle théorique général, fondement de l'analyse, permettant d'établir des relations entre les variables ciblées par (et dans) l'étude envisagée. Elle suppose notamment l'élaboration d'un diagramme sommaire où sont décrits (par des flèches) les effets directs ou indirects de certaines variables (observées ou non) sur d'autres. Elle permet de décomposer le modèle (ou un des modèles) explicatif(s) d'un phénomène particulier, en plusieurs facteurs susceptibles de l'influencer.
- 14 Si on cherche, par exemple, à expliquer un concept A pour lequel on fait l'hypothèse que B et C en sont deux facteurs explicatifs et que l'on suppose de plus que A se décompose en deux dimensions (A' et A''), des hypothèses doivent être posées sur les relations qui lient ces différentes variables théoriques (Kestemont, 1997).

Graphique 1. Schéma d'une théorie principale

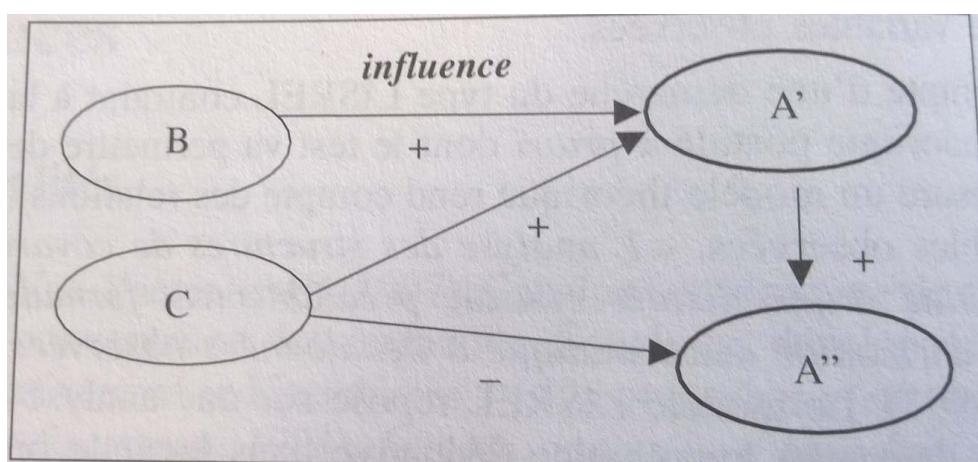


DIAGRAM I. Diagram of one main theory

- 15 L'hypothèse faite repose sur l'idée selon laquelle C a un effet positif sur A', mais négatif sur A'', que B influence uniquement A' (et positivement), et que A' (variable que l'on cherche à expliquer) a un effet positif sur A''. Sont postulés d'une part, les relations qui lient les différents concepts, et d'autre part le sens de ces relations, et leur impact (positif ou négatif).

La spécification d'une théorie auxiliaire ⁴ ou le passage du concept à la mesure (le modèle de mesure)

- 16 Après avoir défini la théorie principale, l'une des premières conditions qui doit être remplie, pour tester l'ajustement du modèle théorique aux données, réside dans la traduction des concepts ou des variables théoriques (hypothétiques qui découlent de la théorie énoncée *a priori*) de la théorie principale en variables observables. Pour cela, les variables latentes vont être opérationnalisées, c'est-à-dire que l'on va procéder à un choix d'indicateurs permettant de rendre compte le plus fidèlement des concepts inobservables. Cette étape consiste à évaluer un ou plusieurs modèles de mesure. Ils vont permettre de spécifier les relations entre indicateurs et concepts mais impliquent toutefois l'existence de problèmes quant à la fiabilité (erreurs de mesure sur les indicateurs) et à la validité des indicateurs choisis (intensité du lien entre concepts et indicateurs). LISREL permettra toutefois de vérifier le degré d'acceptabilité de la mesure et d'estimer les liens que l'on suppose.
- 17 Dans cette structure, A' et A'' sont les variables latentes endogènes, C et B les variables latentes exogènes. Le concept B est mesuré par trois indicateurs ⁵, C et A'' par deux, et A' par quatre. Ainsi les variables observables (symbolisées par des carrés par convention) vont donner ensemble une mesure d'un concept

GRAPHIQUE 2. Représentation de la théorie principale et des théories auxiliaires

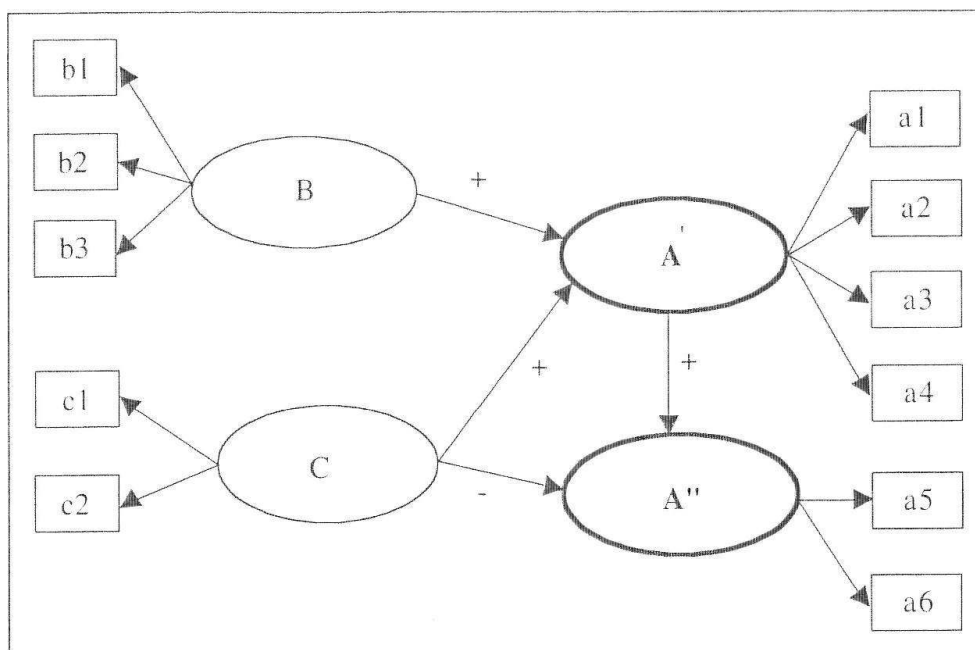


DIAGRAM 2. Representation of the main theory and the auxiliary theories

- 18 non observable directement (représenté par un rond). Nous ferons l'hypothèse qu'un indicateur est associé à une seule variable latente ⁶. On distingue ainsi les modèles de mesure, estimés en premier par LISREL et qui constituent ensemble la théorie auxiliaire. Dans cet exemple, elle s'identifie à l'explication des quatre différents concepts par leurs indicateurs respectifs ; ceci signifie également que chacune des quatre variables latentes se reflète dans ses indicateurs. En seconde étape est identifié le modèle structurel

(relations entre B, C, A'. A'') qui décrit des liens de causes à effets entre les construits ou représentations théoriques (variables latentes).

- 19 Ainsi, la théorie principale, comme les théories auxiliaires, sont basées sur des relations de causalité, et sur l'analyse des chemins de causalité (path analysis), dans laquelle une variable peut avoir un effet direct ou indirect sur une autre (dans l'exemple développé précédemment, il est à noter d'une part un effet direct de C sur A'' et d'autre part un effet indirect de C sur A'' qui transite par A'). Les spécifications de la théorie principale, comme de la théorie auxiliaire, permettent en fait de tester la récursivité des relations entre variables. Dans la théorie auxiliaire, les flèches qui partent des variables latentes et qui vont vers les variables observées correspondent à des saturations. Les relations dans les modèles de mesure (théorie auxiliaire) sont nécessairement récursives, ce qui n'est pas forcément vérifié dans la théorie principale, les flèches ne représentant pas des saturations, mais le poids d'une variable latente sur une autre. Les
- 20 relations entre variables latentes endogènes peuvent ainsi être récursives ou non (existence d'un cycle). En référence à l'exemple développé, un effet de A'' sur A' pourrait être envisagé simultanément à un effet de A' sur A''.

La formalisation en modèles mathématiques

Une fois définies les théories principale et auxiliaires, l'étape suivante consiste à associer, à la structure causale hypothétique, un système d'équations structurelles. En partant d'un diagramme LISREL (graphique 3), différent du diagramme établi précédemment (car supposant certaines conventions d'écriture), plusieurs règles sont à respecter.

GRAPHIQUE 3. Présentation d'un diagramme LISREL (path diagramme) ⁷

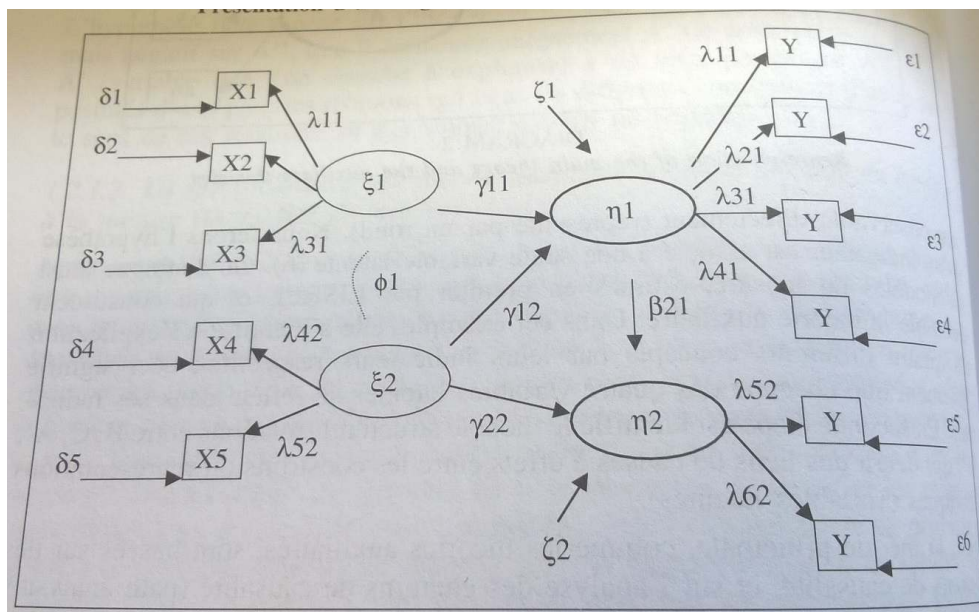


DIAGRAM 3. Presentation of a LISREL diagram

- 21 Pour les variables : LISREL fait la distinction entre variables observées et variables latentes d'une part, entre variables exogènes et endogènes d'autre part. Il distingue également les erreurs sur les équations et les erreurs sur les variables :
- 22 ξ : représente les variables latentes exogènes

- 23 η : les variables latentes endogènes
- 24 χ : les variables observées ou indicateurs des variables latentes exogènes
- 25 Y : les variables observées ou indicateurs des variables latentes endogènes
- 26 δ : les erreurs sur les variables observées X
- 27 ε : les erreurs sur les variables observées Y
- 28 Pour les paramètres : LISREL estime une série de paramètres rendant compte des relations causales formulées *a priori*, soit :
- 29 $\lambda^2(\lambda^y)$ représente l'impact d'une variable latente exogène (respectivement endogène) sur une variable observée
- 30 γ : impact d'une variable latente exogène sur une variable latente endogène β : impact d'une variable latente endogène sur une autre variable latente endogène
- 31 ζ : représente tous les facteurs implicites (non explicités dans la structure causale) qui peuvent avoir un effet spécifique sur les variables latentes endogènes
- 32 Φ : représente les relations entre variables latentes exogènes.
- 33 Dans notre exemple :
- 34 B et C sont des variables latentes exogènes (ξ)
- 35 A' et A'' , des variables latentes endogènes (η)
- 36 b_1, b_2, b_3, c_1, c_2 , les indicateurs des variables latentes exogènes (X)
- 37 $a_1, a_2, a_3, a_4, a_5, a_6$, les indicateurs des variables latentes endogènes (Y).
- 38 La spécification formelle du modèle sous forme d'équations mathématiques dérive du diagramme LISREL, et peut être exprimée par trois séries d'équations matricielles dont on peut distinguer :
- 39 Pour les modèles de mesure :
- 40 $X = \Lambda_x \xi + \theta_\delta$
- 41 $Y = \Lambda_y \eta + \theta_\varepsilon$
- 42 où θ_δ (et θ_ε) est la matrice des erreurs de prédiction pour les indicateurs des variables construites exogènes (respectivement la matrice des erreurs de prédiction des indicateurs des variables construites endogènes).
- 43 Pour le modèle structurel :
- 44 $\eta = B\eta + \Gamma\xi + \Psi$
- 45 où Ψ est la matrice de corrélations des ζ (erreurs des équations structurelles).
- 46 Le détail de quelques-unes de ces relations, amène à:
- 47 $X_1 = \lambda_{11}\xi_1 + \delta_1$
- 48 $X_2 = \lambda_{21}\xi_1 + \delta_2$
- 49 ...
- 50 $Y_1 = \lambda_{11}\eta_1 + \varepsilon_1$
- 51 ...
- 52 $\eta^1 = \gamma^{11}\xi^1 + \lambda^{12}\xi^2 + \xi^1$
- 53 $\eta_2 = \gamma_{22}\xi_2 + \beta_{21}\eta_1 + \zeta_2$

- 54 L'utilisation de modèles du type LISREL suppose que l'on accepte implicitement un certain nombre d'hypothèses, à savoir notamment :
- 55 la linéarité, l'additivité des relations entre variables,
- 56 la non-corrélation des erreurs entre elles
- 57 la non-corrélation des erreurs (préférable dans notre cas) et des variables explicites indépendantes,
- 58 la récursivité des relations (unidirectionnelles) entre indicateurs et variables latentes d'une part, la récursivité ou non des relations entre variables latentes (relations unidirectionnelles ou bidirectionnelles) d'autre part.
- 59 De plus, les hypothèses nécessaires au test de l'inférence causale sont les suivantes :

$$\begin{array}{ll}
 E(\zeta) = 0 & E(\varepsilon, \zeta) = 0 \\
 E(\delta) = 0 & E(\varepsilon, \eta) = 0 \\
 E(\varepsilon) = 0 & E(\delta, \eta) = 0 \\
 E(\zeta, \eta) = 0 & E(\varepsilon, \zeta) = 0 \\
 E(\delta, \xi) = 0 & E(\delta, \zeta) = 0
 \end{array}$$

L'analyse statistique du modèle

- 60 La troisième étape automatisée dans le logiciel consiste à dériver, à partir des équations du modèle, les équations des structures de covariances et les valeurs théoriques des variances et des covariances (des variables observées) impliquées par le modèle (Aish-Van Vaerenbergh, 1997). Cela amène à comparer deux séries de valeurs sur les variables, les valeurs théoriques impliquées par le modèle, et les valeurs observées. Toutefois, pour effectuer cette comparaison, il faut que tous les paramètres du modèle soient identifiés et même sur-identifiés, c'est-à-dire que l'on ait imposé certaines contraintes au modèle, en formulant des hypothèses concernant les relations entre variables observées ou non. Ces hypothèses peuvent se traduire soit par la nullité de certains paramètres, soit par l'imposition de valeurs différentes pour quelques-uns d'entre eux (ces contraintes établissant des restrictions sur la matrice des valeurs théoriques des variances-covariances)⁸.
- 61 L'expression des variances et covariances des variables observées comme fonction des paramètres du modèle est l'hypothèse fondamentale de l'analyse des structures de covariances (modèles LISREL). Elle postule en fait que les variances et covariances observées sont fonction des paramètres du modèle. L'estimation des paramètres libres (ou égaux entre eux sans que leur valeur soit précisée à l'avance) est opérée par le programme (Bacher, 1987).
- 62 LISREL offre plusieurs méthodes d'estimation. Deux d'entre elles (parmi les sept disponibles) sont celle du maximum de vraisemblance lorsque l'on travaille avec des variables continues dont la distribution est normale, et celle des moindres carrés pondérés lorsque l'on s'intéresse à des variables ordinales ou non distribuées

normalement. Chacune d'entre elles vise à minimiser la distance entre valeurs théoriques et valeurs observées.

L'évaluation du modèle

- 63 Lorsque les différents modèles de mesure sont acceptés et validés par LISREL, l'étude de l'ajustement du modèle théorique dans son ensemble (modèles de mesure et modèle structurel) aux observations, doit être envisagée. Elle consiste à savoir si globalement les valeurs théoriques sont en accord avec les valeurs empiriques. « *L'évaluation de l'ajustement empirique se fonde sur la distance entre les valeurs théoriques et les valeurs observées* » (Aish-Van Vaerenbergh, 1997). LISREL donne la valeur de plusieurs indicateurs statistiques qui permettent d'évaluer la distance entre les deux séries de données et de conclure à l'acceptation ou au rejet du modèle causal.
- 64 Les indicateurs d'ajustement calculés par LISREL se scindent en deux groupes distincts, à savoir i) les indicateurs de l'ajustement global du modèle théorique (postulé *a priori*) au modèle empirique et ii) des indicateurs de l'ajustement du modèle au niveau local. La présentation de la spécificité de quelques-uns d'entre eux utiles à la suite de l'analyse, paraît ici nécessaire.

Les indicateurs de l'ajustement global du modèle

L'un des premiers critères qui permet d'évaluer l'ajustement global du modèle théorique aux données, est celui du Chi-deux.

- 65 • *Indice du χ^2* : le test du Chi-deux permet en fait de tester l'hypothèse H_0
- 66 selon laquelle la matrice des covariances observée est générée par le modèle théorique postulé *a priori* contre l'hypothèse alternative H , selon laquelle la matrice des covariances est une matrice sans restriction (sur les paramètres du modèle). Le rejet de l'hypothèse nulle implique que la matrice des covariances basée sur le modèle ne reproduit pas de façon acceptable la matrice des covariances entre variables observées (Godard, 1996). D'aucuns pensent toutefois que le test de cette hypothèse est trop restrictif, dans la mesure où la question posée reste l'adéquation du modèle en termes de compatibilité au sens large (Blanchard *et al.*, 1997).
- 67 De plus, une contrainte existe quant à la validité de cette statistique du χ^2 dans la mesure où, s'il permet de comparer les deux matrices, il reste très sensible à la normalité des variables de mesure, et à la taille de l'échantillon. Le test est valide uniquement avec l'utilisation des covariances, il ne s'applique pas aux corrélations. Jöreskog et Sörbom recommandent alors de l'utiliser comme une mesure de l'ajustement plutôt que comme une épreuve de validation et de le comparer à son degré de liberté, utilisé comme terme de référence pour juger de sa grandeur (Bacher, 1987). Il sera ainsi utile pour comparer deux modèles s'appliquant aux mêmes données, différant uniquement par le nombre de contraintes imposées au modèle (modèles emboîtés) ⁹.
- 68 De façon à pallier les difficultés liées à l'utilisation du χ^2 , Jöreskog et Sörbom ont proposé d'autres indicateurs (G.F.I., A.G.F.I.) indépendants de l'effectif et du nombre de paramètres du modèle, permettant d'apprécier l'adéquation du modèle théorique aux données. Ces indicateurs sont ainsi couramment employés pour évaluer l'ajustement global d'un modèle théorique.

- 69 *Indices G.F.L (Goodness of Fit Index) et A.G.F.I. (Adjusted Goodness of Fit Index)*: indiquent la proportion de la matrice de variances-covariances des variables observées (S) qui peut être expliquée par le modèle théorique (par E : matrice des variances-covariances calculée à partir du modèle). Ces indices doivent être le plus près de 1. Une valeur de 0,9 est généralement considérée comme satisfaisante. Bender et Bonnet (1980) ont développé d'autres indices : N.F.I. (Normed Fit Index), P.N.F.I. (Parsimony Normed Fit Index), C.F.I. (Comparative Fit Index...). Les deux d'entre eux les plus utilisés sont détaillés par la suite.
- 70 *Indice N.F.I. (Normed Fit Index) et indice N.N.F.I. (No Normed Fit Index)*: sont des indicateurs basés sur l'écart à un modèle non explicatif. L'indice N.F.I. mesure l'accroissement d'ajustement entre deux modèles emboîtés appliqués aux mêmes données par rapport à l'ajustement que l'on obtiendrait avec un modèle nul contraint au maximum. Une valeur proche de 1 témoigne du pouvoir explicatif du modèle. Toutefois, la justification de l'utilisation de ces indices est controversée (problème de définition du modèle contraint au maximum notamment).
- 71 Des indicateurs (R.M.R., R.M.S.E.A.) relatifs aux résidus (écarts entre les deux matrices) permettent également de tester l'ajustement global du modèle théorique aux données empiriques.
- 72 *Indice R.M.R. (Root Mean Square Residuals)*: indique la valeur moyenne des résidus du modèle. C'est une mesure de la moyenne des variances et covariances résiduelles. Sa valeur, comprise entre 0 et 1, doit être la plus petite possible.
- 73 *Indice R.M.S.E.A. (Root Mean Square Error of Approximation)*: se rapporte à la significativité des résidus, et à l'erreur d'approximation. L'approximation est acceptable si la valeur du R.M.S.E.A. est inférieure ou égale à 0,05.
- 74 D'autres indicateurs existent et sont donnés par LISREL. Seuls les plus fréquemment employés ont été repris ici.

Les Indicateurs d'ajustement local

- 75 Une fois réalisée, l'évaluation de l'ajustement du modèle dans son ensemble, l'examen et l'interprétation des paramètres estimés peuvent être effectués. Pour interpréter les paramètres estimés, l'indicateur dont on dispose est le *t* de Student.
- 76 • *t de Student* : il permet de déterminer si les paramètres estimés du modèle sont significativement différents de 0, et d'estimer les risques que les coefficients soient du signe opposé à celui mis en évidence par le modèle. Pour être significativement différent de 0, la valeur du *t* doit être supérieure ou égale à 1,96 (valeur au seuil de 5 %). Si tel n'est pas le cas ($t < 1,96$), on peut penser que la relation entre les deux variables n'est pas significative, le lien peut être éliminé. Un modèle plus restrictif est alors obtenu.
- 77 D'autres indicateurs permettent d'évaluer l'ajustement local du modèle, tels les résidus standardisés (Standardized Residuals) qui mesurent les covariances résiduelles entre deux indicateurs. La mesure la plus utilisée reste le *t* Student.

Exemple d'application et apports des modèles en structure de covariances : la mise en évidence de compétences transversales pour expliquer l'insertion au collège

- 78 Un exemple d'application de l'analyse en structure de covariances et tiré d'une étude récente (Morlaix, 1999), s'intéresse à l'efficacité du processus d'apprentissage et plus particulièrement aux effets de la structure d'allocation de temps à l'école primaire (C.M.2) sur la qualité de l'insertion des élèves au collège ¹⁰. De façon à isoler quels pouvaient être les arbitrages entre les différentes composantes de l'utilisation du temps en C.M.2 qui favoriseraient l'insertion au collège des élèves les plus en difficulté, un modèle théorique a été postulé *ci priori* et testé par l'utilisation de LISREL. Le modèle initialement postulé reposait sur les hypothèses suivantes :
- 79 D'une part, a été faite la supposition selon laquelle l'acquisition par les élèves de compétences transversales pendant l'année de C.M.2 ¹¹ jouait un rôle sur l'insertion au collège des élèves. Ces compétences (savoir rédiger, savoir analyser, savoir s'exprimer,...) ne peuvent être observées directement et constituent des variables latentes (exogènes) pour lesquelles on a cherché à estimer l'effet sur une autre variable latente (endogène cette fois) car inobservée : l'insertion au collège.
- 80 Une fois, la théorie principale du modèle formulée, les modèles auxiliaires ont été postulés. On a ainsi supposé que les indicateurs de compétences transversales acquises pouvaient être les scores obtenus par les élèves à une évaluation nationale à l'entrée en 6e (cette évaluation étant composée de neuf grands groupes d'items : cinq en mathématiques et quatre en français). Ce sont les neuf scores obtenus qui seraient le reflet de compétences transversales acquises en fin de primaire. Les indicateurs de l'insertion au collège sont quelque peu différents puisqu'ils reposent d'une part sur les résultats obtenus par les élèves à une évaluation de fin de 6e, ainsi que sur l'âge des élèves à l'entrée au collège.

GRAPHIQUE 4. Modèle théorique mettant en relation les scores obtenus aux différents items, la production de transversales et l'insertion des élèves au collège

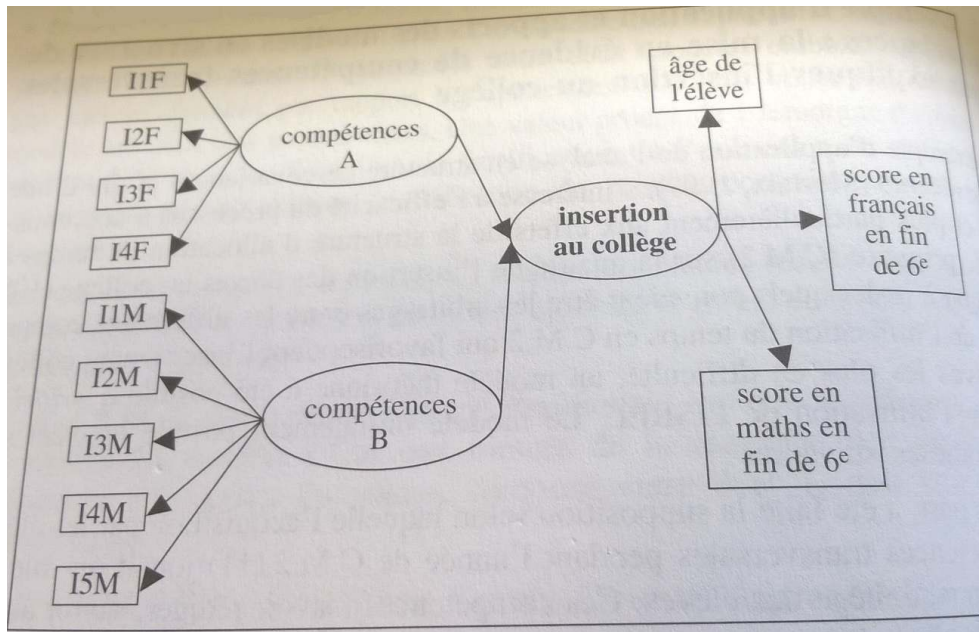


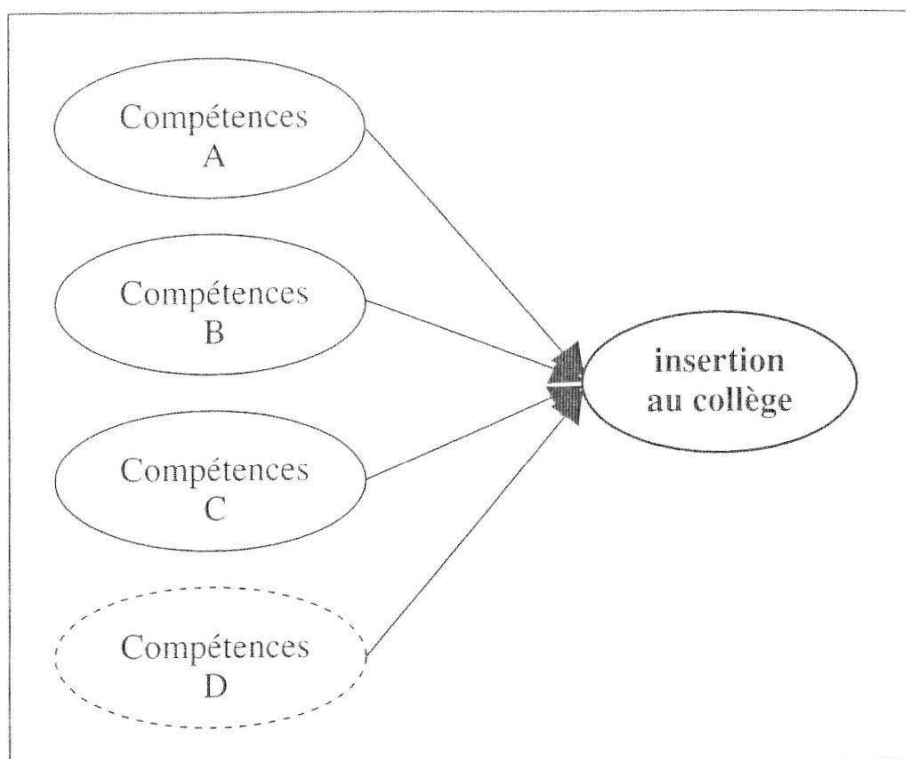
DIAGRAM 4. Ideal model linking the scores obtained for different items

- 81 Ce modèle peut être testé en reprenant les différentes étapes présentées dans le chapitre précédent.

La théorie principale du modèle

- 82 La théorie principale permet de prendre en compte et de mettre à jour les relations que l'on suppose exister entre variables latentes (exogènes et endogènes du modèle). Elle définit le sens et la récursivité des relations. Initialement, la théorie principale du modèle était la suivante :

GRAPHIQUE 5. Théorie principale du modèle initial

DIAGRAM 5. *Principal theory of the initial model*

- 83 L'une des hypothèses formulées consiste à supposer l'existence d'un effet des aptitudes transversales acquises (le nombre des compétences A, B, C, D et leur nature étant à définir) sur l'insertion des élèves au niveau éducatif supérieur. Les variables latentes caractérisant les compétences transversales sont des variables exogènes dont on ne cherchera pas à expliquer la variabilité. À l'inverse, on cherchera à expliquer la part de variance de la variable latente endogène (insertion) par la prise en compte des différentes variables latentes exogènes (compétences). Les relations entre variables sont toutes présumées récursives et positives, et l'on suppose que la maîtrise des différentes compétences transversales joue positivement, sur l'insertion des élèves, l'année suivante. Les relations entre variables latentes étant définies, les relations entre ces variables et leurs indicateurs (théorie auxiliaire) doivent être déterminées.

La théorie auxiliaire du modèle initialement formulé

Identification et détermination du nombre des aptitudes transversales (variables exogènes)

- 84 La théorie auxiliaire permet de déterminer les relations liant indicateurs et variables latentes. Comme il est précisé précédemment, chacun des indicateurs choisis doit être une mesure fiable et valide ¹² de la variable latente à laquelle
- 85 il est associé. La détermination du nombre de variables exogènes (compétences) et leur identification passe par l'étude des modèles de mesure liant chacune des variables latentes à ses indicateurs. Les indicateurs des variables latentes exogènes (compétences

transversales) sont les scores obtenus aux cinq items de mathématiques, et ceux obtenus aux quatre items de français en fin de C.M.2. Trois variables latentes exogènes ont alors été mises à jour, et les relations mises en évidence entre les variables latentes exogènes et leurs indicateurs sont les suivantes (*graphique 6*).

GRAPHIQUE 6. Modèles de mesure des construits exogènes

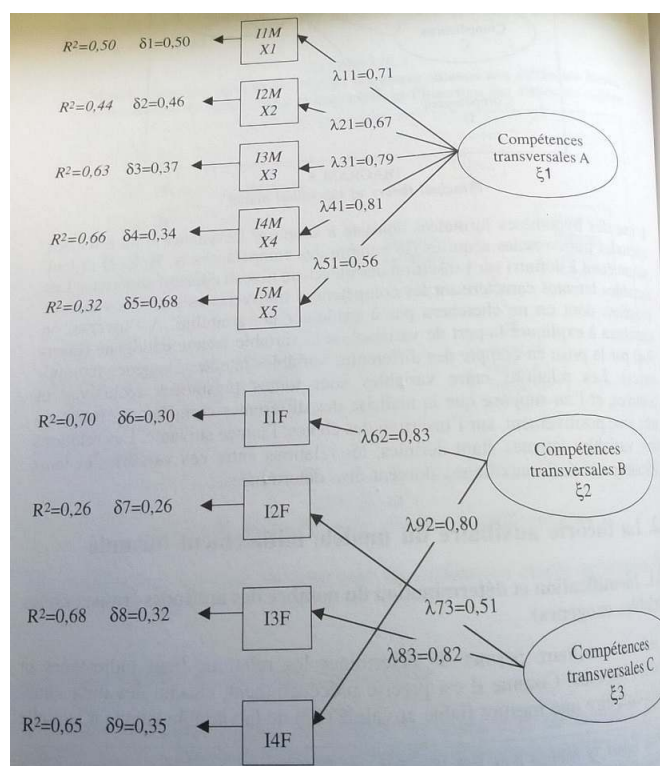


DIAGRAM 6. Models of the measurement of external factors of the value construct

- 86 Trois variables latentes exogènes (groupes d'aptitudes) sont ainsi mises à jour. De façon à pouvoir identifier plus précisément les compétences A, B, et C, plusieurs constats peuvent être faits.
- 87 La variable latente exogène ξ1 commune aux cinq items de maths pourrait représenter des compétences de type scientifique ou logique (savoir analyser un problème, savoir raisonner,...). C'est pourquoi, ces compétences mesurées conjointement par les cinq items de mathématiques, seront désignées par la suite comme les « aptitudes logiques ».
- 88 Concernant les items de français, à savoir l'item 1 (compréhension) et l'item 4 (outils lexicaux et grammaticaux) d'une part ; l'item 2 (production de textes) et l'item 3 (outils en production de textes) d'autre part : les deux premiers items (1 et 4) mesurent ensemble une variable latente exogène ξ2. L'hypothèse sous-jacente consiste à penser que cette variable latente exogène est une variable caractérisant des compétences acquises dans le domaine des attitudes, et de l'apprentissage de la vie sociale, témoignant du fait que l'élève comprend une situation donnée, et sait appliquer des règles, les deux items (1 et 4) pouvant représenter en effet des tâches d'exécution (comprendre un énoncé, appliquer et utiliser des règles, des outils lexicaux et grammaticaux...). Dans cette optique, les compétences transversales B (ξ2) seront qualifiées par la suite « d'aptitudes d'exécution ».

- 89 La dernière variable latente exogène C (ou ξ_3) est représentée par les items de français (items 2 et 3). Cette aptitude peut s'apparenter à une compétence représentant la capacité de l'élève à produire un texte. On peut penser par ailleurs que ces aptitudes font appel à un « savoir inventer », ou un « s'avoir innover », qui traduit la créativité de l'élève et une certaine autonomie. C'est pourquoi, ces compétences seront nommées par la suite « compétences d'innovation ».
- 90 Il paraît toutefois utile de rappeler, après la définition précise des trois variables latentes exogènes, que l'appellation et la détermination de ces variables latentes incombent au seul chercheur qui, s'appuyant sur la théorie, détermine de façon subjective et arbitraire les variables latentes qui lui paraissent les plus pertinentes en fonction de ce qu'il cherche à démontrer.
- 91 De la même façon, le modèle relatif à la variable latente endogène (insertion au collège) a été analysé.

Modèle de mesure de la variable latente endogène

- 92 La variable latente endogène supposée être expliquée par les variables latentes exogènes mises en évidence précédemment (aptitudes logiques, aptitudes d'innovation, aptitudes d'exécution), est également mesurée par un ensemble d'indicateurs. Ces indicateurs mesurant l'insertion sont à rapprocher des scores obtenus par les élèves en français et en maths en fin d'année de 6e et par l'âge des élèves à leur entrée au collège. Les indicateurs retenus pour mesurer l'insertion de l'élève au collège sont finalement au nombre de trois et sont supposés constitués des indicateurs valides de l'intégration des élèves ; ceci dans la mesure où, comme le constate déjà Caille (1997), « c'est toujours au niveau d'acquisitions en français et en mathématiques et à l'âge de l'élève à l'entrée en 6e que sont associées les disparités de réussite les plus tranchées ». Le modèle de mesure de la variable latente endogène permet de mettre en évidence, les relations existantes entre la variable latente endogène et ses indicateurs, soit :

GRAPHIQUE 7. Modèle de mesure du construit endogène

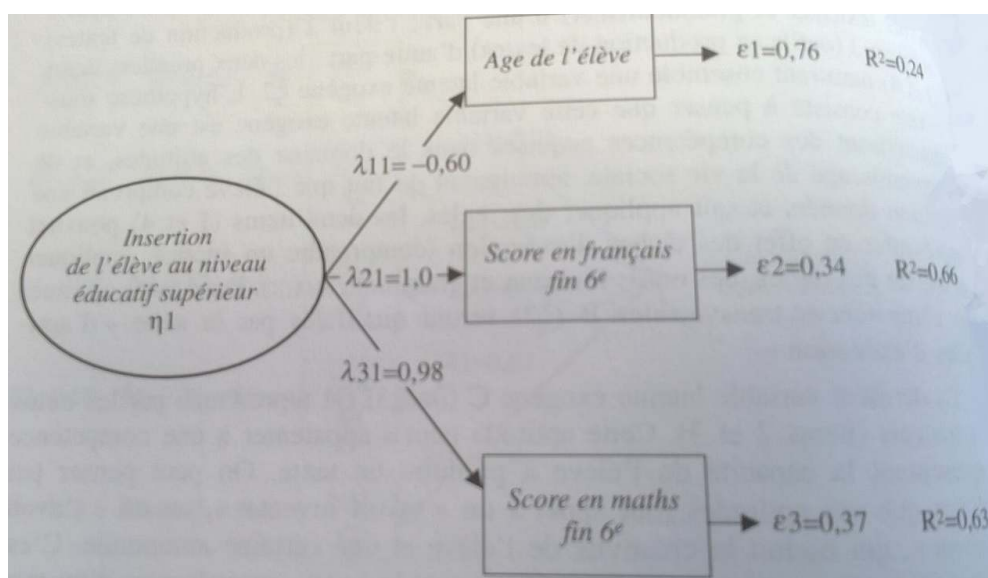


DIAGRAM 7. Model of the measurement of the internal factors of the value construct

À la lecture du graphique, une relation négative lie la variable « âge de l'élève » et la variable latente « insertion de l'élève », ce qui signifie que plus l'élève est âgé en 6e, moins son insertion au collège sera facile. De la même façon, une relation positive est mise à jour entre les scores en français et en maths en fin de 6e et l'insertion des élèves. Il est toutefois à noter la restriction imposée au modèle en fixant à 1 le paramètre 121, de façon à imposer l'unité de mesure de la variable latente endogène (de la même façon, $\varepsilon_3 = 0,37$ $R^2 = 0,63$, on aurait pu choisir $\lambda_{31} = 1$). L'ensemble des coefficients liant la variable latente endogène à ses indicateurs sont significatifs au seuil de 1 %. Le modèle structurel finalement retenu peut alors être présenté.

Le modèle structurel retenu

93 Le modèle théorique global prend la forme suivante :

GRAPHIQUE 8. Modèle structurel retenu

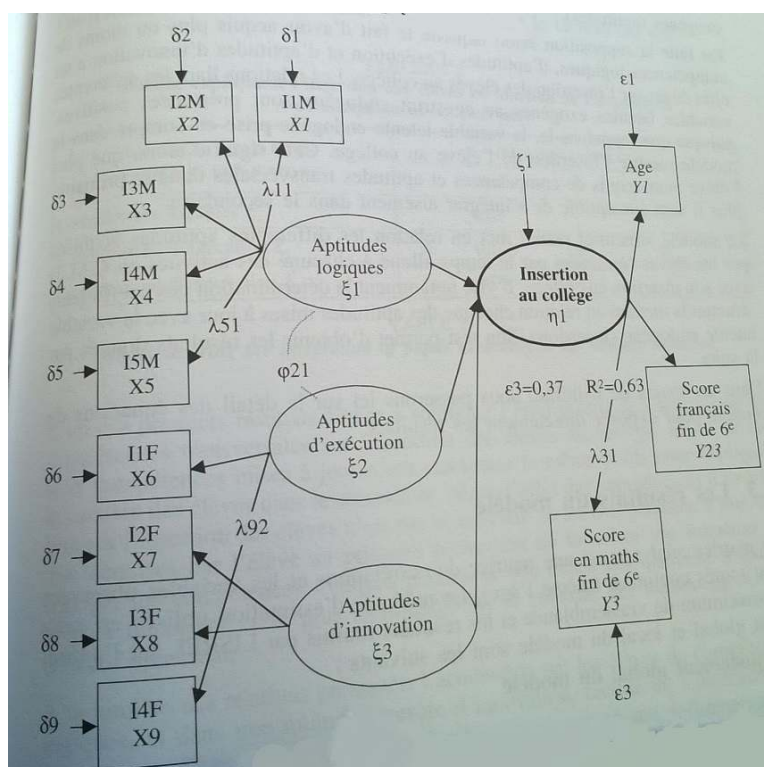


DIAGRAM 8. Structural model selected

94 Par souci de lisibilité, seulement quelques-uns des paramètres λ , témoignant de la relation entre un indicateur et une variable latente, ont été représentés sur le graphique. Comme il a été établi précédemment, η représente la variable latente endogène (insertion des élèves) et les ξ les variables latentes exogènes (aptitudes). Les Y (âge de l'élève, scores en français et en maths en fin de 6e) sont les indicateurs de la variable latente endogène, tandis que les X (scores obtenus par les élèves aux différents items de français et de maths à l'entrée en 6e) sont les indicateurs des variables latentes exogènes. Les λ caractérisent donc les relations entre les variables latentes et leurs indicateurs, les γ les relations entre variables latentes exogènes (aptitudes) et la variable latente **endogène** (insertion). Enfin, les δ représentent les erreurs sur les indicateurs des exogènes, et les ε les erreurs sur les

indicateurs de la variable latente endogène. ζ représente les facteurs implicites qui peuvent avoir un effet sur la variable latente endogène et γ les relations entre variables latentes exogènes (aptitudes).

- 95 Est faite la supposition selon laquelle le fait d'avoir acquis plus ou moins de compétences logiques, d'aptitudes d'exécution et d'aptitudes d'innovation a un effet direct sur l'insertion des élèves au collège. Les relations liant les différentes variables latentes exogènes au construit endogène sont présumées positives, puisque que rappelons-le, la variable latente endogène prise en compte dans le modèle mesure l'insertion de l'élève au collège. Cela signifie alors, que plus l'élève aura acquis de compétences et aptitudes transversales dans le primaire, plus il sera susceptible de s'intégrer aisément dans le secondaire.
- 96 Le modèle structurel retenu met en relation les différentes aptitudes acquises par les élèves (mesurées par le temps alloué à chacune des activités en C.M.2) avec son insertion au collège. Il vise notamment la détermination des coefficients structurels mettant en relation chacune des aptitudes mises à jour avec la variable latente endogène (insertion). Son test permet d'obtenir les résultats donnés par la suite.
- 97 Dans un souci de lisibilité, nous passerons ici sur le détail des équations du modèle pour exposer directement les résultats.

Les résultats du modèle

- 98 La matrice analysée est une matrice de corrélations et les variables observées sont toutes continues (même l'âge). La méthode d'estimation utilisée est celle du maximum de vraisemblance et les résultats fournis par LISREL sur l'ajustement global et local du modèle sont les suivants :
- Ajustement global du modèle

TABLEAU 1. Indices d'ajustement global du modèle

DDL	χ^2	GFI	AGFI	NFI	NNFI	RMSEA	RMR
45	118	0,97	0,95	0,97	0,97	0,049	0,026

TABLE 1. Global adjustment of the model

- 99 Les indices d'ajustement global du modèle témoignent de son acceptabilité. Certes, le modèle ne s'ajuste pas parfaitement, mais les valeurs des indicateurs sont dans l'ensemble assez satisfaisantes (G.F.I., A.G.F.I., N.F.I., N.N.F.I. étant supérieurs à 0,9, et les indices R.M.S.E.A. et R.M.R. attestant de valeurs inférieures à 0,05).
- Ajustement local du modèle :

- 100 La relation testée par ce modèle prend la forme :

101 $\eta_1 = f(\xi_1, \xi_2, \xi_3)$

- 102 Les résultats de cette modélisation sont donnés dans le tableau qui suit :

TABLEAU 2. Modèle expliquant l'insertion des élèves en fonction de l'acquisition de compétences transversales

VARIABLES À EXPLIQUER VARIABLES EXPLICATIVES	INSERTION AU COLLÈGE	
	coeff.	t
Compétences logiques ξ_1	0,63	***
Compétences d'exécution ξ_2	0,10	ns
Compétences d'innovation ξ_3	0,08	*

TABLE 2. Model showing the integration of pupils according to the abilities acquired

- 103 L'un des premiers résultats amène à souligner l'effet positif de l'acquisition de compétences transversales sur l'insertion des élèves au collège, même si les trois compétences mises à jour n'ont pas toutes le même poids pour expliquer la carrière des élèves dans le secondaire. Ainsi, l'effet des compétences d'exécution sur l'insertion des élèves n'est pas significatif. Si les compétences d'innovation acquises par l'élève au primaire permettent de favoriser son insertion au collège, les compétences logiques paraissent être celles qui exercent le poids le plus important et le plus significatif sur la réussite des élèves au niveau éducatif supérieur.
- 104 À la lumière des résultats présentés, l'acquisition par les élèves de compétences logiques, et dans une moindre mesure d'innovation, faciliterait l'insertion des élèves au collège. On pourrait ainsi s'interroger sur ce qui influencerait de façon significative le développement de ces compétences au cours du primaire, notamment durant l'année de C.M.2.
- 105 Plus particulièrement, il paraît intéressant de déterminer quels sont les temps en C.M.2 qui influent significativement sur l'acquisition par les élèves, de compétences logiques et d'innovation, ces deux groupes de compétences étant déterminants pour la bonne insertion au collège. Il est possible, pour ces deux groupes (mesurés par les résultats obtenus à différents items), d'isoler les scores qui les reflètent le mieux. Une fois les scores aux items qui reflètent le mieux les compétences logiques et d'innovation acquises, identifiés, il sera possible de trouver quels sont les temps en C.M.2 qui influent le plus sur les scores obtenus. Le schéma suivant résume l'ensemble de la démarche suivie dans cette étude :

GRAPHIQUE 9. Schéma général de la modélisation envisagée

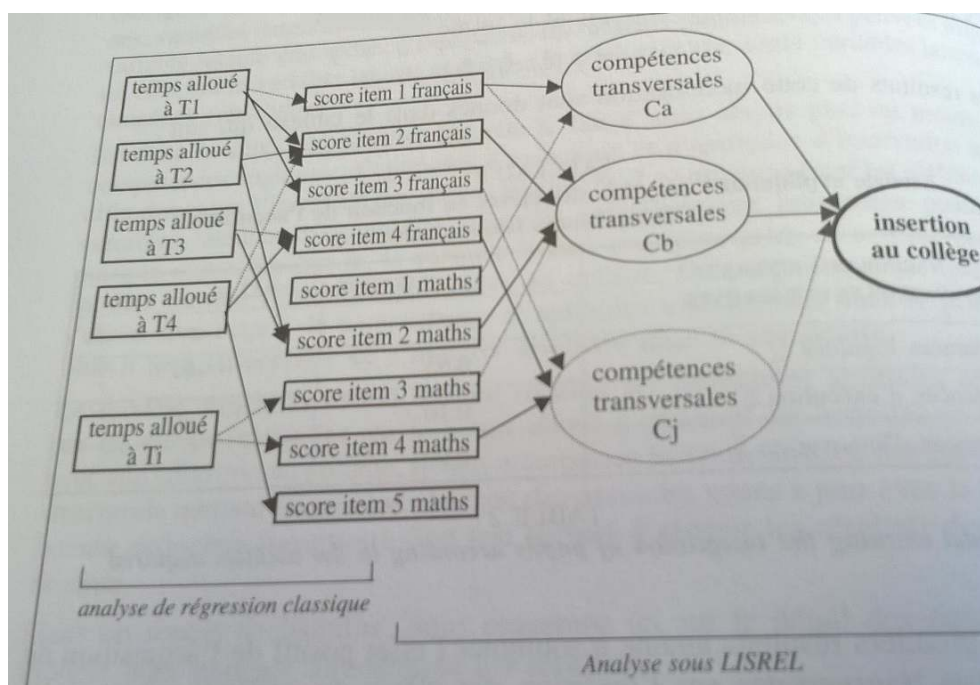


DIAGRAM 9. General outline of considered process

- 106 Une fois trouvés les temps en C.M.2 qui influent le plus sur les scores obtenus, et par là sur l'acquisition des compétences logiques et d'innovation, donc sur l'insertion au collège, il est possible d'instruire des arbitrages dans les choix à réaliser entre les différentes composantes du temps en primaire de façon à favoriser l'insertion des élèves au collège. Cette problématique n'étant pas l'objet direct de cette étude, nous ne détaillerons pas, par la suite, l'ensemble de la démarche permettant d'éclairer les arbitrages à réaliser, l'essentiel de la démarche exposée étant de mettre à jour l'intérêt de l'utilisation de modèles tels que ceux développés par LISREL.

Conclusion

- 107 L'analyse sous LISREL comporte de nombreuses limites dont l'une des principales tient à la complexité de compréhension et d'utilisation des modèles visés. La nécessité de faire de multiples choix apparaît sans cesse dans ce type d'analyse, et de nombreuses limites (telles que la dénomination des variables latentes) nécessitent de disposer d'un modèle théorique fort. Malgré toutes ces limites, l'analyse de ces modèles en structure de covariances permet notamment de répondre aux besoins de chercheurs qui souhaitent parvenir à des systèmes explicatifs aussi complets que possibles du comportement (Bacher, 1988). Elle présente ainsi des pistes de réflexion nouvelles ouvrant la voie à des travaux enrichis par une démarche méthodologique différente.

BIBLIOGRAPHY

- Achouche, M., & Figari, M. (1996). Observation des performances et estimation des compétences des élèves à l'entrée de la scolarité obligatoire. In *Mesure et évaluation en Éducation*, 19, 3, 5-39.
- Aish-Van Vaerenbergh, A.-M. (1997). « Modèles statistiques et inférences causales : analyse des structures de covariances avec LISREL ». In *Faut-il chercher aux causes une raison ? L'explication causale en sciences humaines*, Aish-Vaerenbergh A.-M. et al., collection sciences-histoire-philosophie (pp. 106-130). Librairie philosophique Vrin. Institut interdisciplinaire d'études épistémologiques.
- Bacher, F. (1987). Les modèles structuraux en psychologie. Présentation d'un modèle LISREL. Première partie. *Le travail humain*, 50, 4, 347-370.
- Bentler, P. M., & Bonnet, D. G. (1980). « Multivariate Analysis with Latent Variables : Causal Modeling », *Annual Review of Psychology*, 31, 419-456.
- Blanchard, S., Vrignaud, P., Lallemand, N., Dosnon, O., & Wach, M. (1997). Validation de l'échelle de motivation en éducation auprès de lycéens français. *L'Orientation Scolaire et Professionnelle*, 26, 1, 33-56.
- Godard, L. (1996). *Conseil d'administration, stratégie et performance financière*, Thèse de doctorat en sciences de Gestion, mention finance. Dijon : Université de Bourgogne.
- Hoyle, R. H. (1995). *Structural Equations Modeling : Concepts, Issues and Applications*. London : Sage publications.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1987). *Causal Analysis : Assumptions, Models and Data. Studying organisations, innovations in Methodology, 1*. London : Sage Publications.
- Jöreskog, K., & Van Thillo (1973). « LISREL : A General Vcomputer Program for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables ». Research Report 73-5. Uppsala (Sweden) : Departement of Statistics, Uppsala University.
- Jöreskog, K., & Sörbom, D. (1982). Recent Developments in Structural Equation Modeling. *Journal of Marketing Research*, 19, 404-416.
- Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the Simplis Command Language*. Chicago : Scientific Software International.
- Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: User's Reference Guide*. Chicago : Scientific Software International.
- Keeves, J. P., & Tuijnman, A. C. (1997). Path analysis and Linear Structural Relations Analysis. In *Educational Research, Methodology and Measurement : An International Handbook*, Keeves J. P. Eds., second edition. Oxford : Pergamon Press and Elsevier Science.
- Kestemont, M.-P. (1997). L'introduction à LISREL. L'apport des modèles à variables latentes à l'analyse des données d'enquête. École d'été du C.N.R.S., C.R.A.P.S.-I.E.P. Lille, 3- 10 septembre 1997.

Morlaix, S. (1999). *Essai sur l'allocation et l'optimisation du temps scolaire. La transition entre école primaire et collège*. Thèse de Doctorat en science économique, Université de Bourgogne, 378 p., janvier 1999.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Dean Stilwell, C. (1989). « Evaluation of Goodness of Fit Indices for Structural Equation Models ». *Psychological Bulletin*, 105, 3, 430-445.

Valette-Florence, P. (1988). Spécificités et apports des méthodes d'analyse multivariée de la deuxième génération. *Recherche et applications en marketing*, 3, 4, 23-56.

Valette-Florence, P. (1993). Dix années de modèles d'équations structurelles : un état de l'art. *Série Recherche, C.E.R.A.G.*, 9-11 mars 1993.

NOTES

1. Les données analysées sont habituellement les variances et les covariances ou les corrélations des variables observées, d'où le terme d'analyse en structure de covariances » (Bacher, 1987).
2. LISREL : Linear Structurel RELationship.
3. Par opposition aux analyses classiques plus traditionnelles, telles que l'analyse factorielle, la régression multiple, l'analyse de variances,... qui peuvent présenter l'une ou l'autre des spécificités énoncées sans les posséder conjointement.
4. Le concept de théorie auxiliaire est dû à Costner et Blabock (1968) cités par Aish-Van Vaerenbergh (1997).
5. D'après Valette-Florence (1988), la nature de la relation entre variable latente et indicateurs peut prendre différentes formes : symétrique (pas de présomption de relations unidirectionnelles entre les indicateurs et les variables latentes), réflexive (on suppose dans ce cas que ce sont les variables latentes qui donnent naissance à quelque chose qui est observé). formative (les variables non observables peuvent être conçues comme des effets et non des causes).
6. Dans le cas contraire, si on conçoit par exemple, que a_4 est un indicateur de A' et A'' , cela suppose implicitement que l'on fasse l'hypothèse que c'est à la fois un indicateur de la cause et de l'effet, ce qui paraît délicat pour mener une analyse de ce type.
7. Ainsi, X constitue une mesure de la validité de l'indicateur considéré. δ , et ϵ des mesures de la fiabilité des indicateurs (respectivement X et Y).
8. Soit t si on considère une matrice à $(p + q)$ variables et le t le nombre de paramètres libres à estimer, le modèle sera sur-identifié si et seulement si $t < ((p + q)(p + q + 1)/2)$.
9. Si par exemple, on supprime une contrainte en libérant un des paramètres précédemment fixé, l'ajustement sera généralement amélioré, mais on aura perdu un degré de liberté. « Dans la mesure où la baisse du X est importante par rapport à la différence du degré de liberté, le gain est sans doute réel » (Baehrer, (1987).
10. Dans cette étude, un échantillon d'environ 1 000 élèves de C.M.2 a été constitué. Des données sur leurs caractéristiques socio-démographiques ont été collectées, ainsi que des renseignements sur les caractéristiques des enseignants de primaire et sur l'utilisation du temps dans la classe. Par la suite, ces élèves ont été suivis en classe de 6e où des informations ont été recueillies sur le contexte classe, l'établissement fréquenté et sur les emplois du temps. Par ailleurs, une évaluation initiale (début de C.M.2), intermédiaire (entrée au collège) et finale (fin de 6e) ont permis d'étudier la progression des élèves en fonction des différentes informations recueillies.
11. On a supposé que ce sont les temps alloués à chacune des activités en C.M.2 qui contribuaient à produire outre des connaissances spécifiques à chacune des disciplines et facilement

mesurables par un test, d'autres compétences transversales à chacune des activités composant le programme et par nature inobservables.

12. Pour rappel, la fiabilité d'un indicateur fait référence aux erreurs de mesure existantes sur cet indicateur, la validité d'un indicateur à l'intensité de la relation entre l'indicateur et le concept qu'il est sensé représenter.

ABSTRACTS

The purpose of this article is modest. First, it aims at showing in a simple way the concepts that make up the structure of latent structural relation models and hereafter their usefulness in the field of research for social sciences. This theoretical approach is then followed by an illustration showing the integration of the way school time is managed within the context of the educational system.

L'ambition de cet article est modeste. Il vise dans un premier temps à présenter de façon simplifiée les notions qui sous-tendent l'utilisation de modèles type LISREL et montrer ainsi leur utilité dans le domaine de la recherche en sciences sociales. Il propose par la suite d'illustrer les propos théoriques énoncés par un exemple d'application portant sur l'utilisation du temps scolaire dans le système éducatif.

INDEX

Mots-clés: Analyse en structure de covariances, LISREL, variables latentes, système éducatif

Keywords: Linear Structural Relationship, LISREL, Educational System

AUTHOR

SOPHIE MORLAIX

est Maître de Conférences à Paris XII. Attachée à l'I.R.E.D.U./C.N.R.S. Thèse de doctorat portant sur l'optimisation du temps scolaire, soutenue en 1999. Travaux de recherche sur l'efficacité et l'équité des systèmes éducatifs et des pratiques enseignantes. I.R.E.D.U./C.N.R.S., 9, avenue Alain Savary, B.P. 47870, 21078 Dijon Cedex. Tél. :03.80.39.54.50. Courriel : sophica@u-bourgogne.fr ou sophie.morlaix@wanadoo.fr